

## 智能多语种新闻采编系统的多模态数据融合研究

李满江<sup>1</sup> 鞠传森<sup>2</sup> 任 鹏<sup>3</sup>  
LI Manjiang JU Chuansen REN Peng

## 摘 要

探讨智能多语种新闻采编系统的多模态数据融合技术, 提出有效的融合方法, 提升系统的智能化水平。对多模态数据进行采集、处理、分析和融合技术研究。多模态数据融合技术可应用于自动化新闻生成与增强、事实核查与假新闻检测、情感分析与舆论监控等场景。未来, 它还可能更智能的自动化采编、基于脑机接口的新闻体验、跨时空新闻场景重建等方面发挥作用。多模态数据融合对新闻采编至关重要, 它不仅丰富了新闻的表达形式, 还提高了内容质量和编辑效率。未来, 多模态融合技术将持续推动新闻行业迈向更高层次的智能化与个性化服务。

## 关键词

多模态数据融合; 新闻采编系统; 新闻信息传播; 情绪响应新闻

doi: 10.3969/j.issn.1672-9528.2024.08.047

## 0 引言

随着全球化和信息技术的飞速发展, 新闻信息的传播已经突破了地域和语言的限制, 呈现出多语种、多模态的特点。传统的新闻采编系统主要依赖于单一的文本数据, 难以全面、准确地捕捉新闻事件的复杂性和多样性。智能多语种新闻采编系统旨在通过引入多模态数据融合技术, 实现对多语种新闻信息的全面采集、处理和分析。多模态数据融合技术能够将来自不同来源、不同形式的数据(如文本、图像、音频、视频等)进行有机融合, 从而提取出更丰富、更全面的信息。在新闻采编领域, 多模态数据融合技术可以帮助系统更准确地识别新闻事件的关键信息, 提高新闻采编的效率和准确性, 形成协同效应, 成为增强新闻报道全面性、真实性和吸引力的关键<sup>[1]</sup>。

多模态数据融合技术在智能多语种新闻采编系统中的应用仍面临诸多挑战, 多模态数据之间的异构性和复杂性使得数据融合过程变得复杂而繁琐。首先要解决的是利用多模态数据中的关键信息, 提高新闻采编的智能化水平。

本研究旨在深入探讨智能多语种新闻采编系统的多模态数据融合技术, 通过对多模态数据的采集、处理、分析和融

合技术的研究, 提出有效的多模态数据融合方法, 实现对多语种新闻信息的全面、准确、高效处理, 有助于提升新闻采编系统的智能化水平。

## 1 新闻采编中多模态数据

在新闻采编中, 常见的多模态数据主要包括以下几种。

(1) 文本数据: 这是新闻报道中最基本的形式, 包括文章正文、标题、摘要、引用语、评论、社交媒体帖子等。文本数据传达了新闻事件的核心信息、背景、细节以及作者的观点和态度。

(2) 图像数据: 包括新闻照片、图表、信息图、漫画等视觉元素。图像能够直观展示新闻现场、人物、事件影响等, 增强新闻的真实感和吸引力。

(3) 视频数据: 视频片段或直播为新闻报道提供了动态视角, 可以展现事件发生的实际过程、采访实录、现场情况等, 增加报道的生动性和即时性。

(4) 音频数据: 如现场录音、采访音频、播客新闻等。音频数据能够捕捉到声音的氛围, 如现场的嘈杂声、采访对象的语气变化, 增强报道的情感表达和真实感。

(5) 社交媒体数据: 从微博、推特、Facebook 等社交平台上收集的数据, 包括文字、图片、视频分享、用户互动(点赞、评论、转发)等, 反映了公众对新闻事件的反应和社会情绪。

(6) 地理位置数据: GPS 坐标、地图信息等, 用于标示新闻事件发生的地点, 或者在交互式新闻中提供事件地理位置的上下文信息。

1. 潍坊北大青鸟华光照排有限公司 山东潍坊 261061

2. 大众报业集团(大众日报社) 山东济南 250014

3. 潍坊日报社 山东潍坊 261000

[基金项目] 山东省重点研发计划山东省科技型中小企业创新能力提升工程项目“智能多语种新闻采编系统”(2022TSGC2368)

(7) 传感器数据：在某些情况下，新闻采编也会利用物联网设备提供的数据，如环境监测站的气候数据、地震仪的地震数据，这些数据可以为自然灾害报道等提供科学依据<sup>[2]</sup>。

(8) 时间序列数据：如股市波动、天气变化等随时间变化的数据，常以图表形式呈现，帮助读者理解事件的发展趋势。

利用这些数据整理出的有效数据类型，主要包括以下几种感知通道的信息类型。

(1) 视觉信息：如图片、视频等，它们是新闻报道中常用的视觉元素，能够直观地展示事件现场或人物形象。

(2) 文本信息：包括新闻文章、标题、摘要等，是传递新闻内容的主要方式。

(3) 听觉信息：听觉信息是通过耳朵获取的，包括声音、语音、音乐、音效等。在新闻采编中，音频数据如采访录音、新闻播报等提供了重要的听觉信息，使读者能够听到新闻人物的原始声音和语调。

(4) 符号信息：如图表、符号、表情等，它们可以辅助文本，以更直观的方式展示数据或情感。

(5) 数值信息：统计数据、调查结果等数值信息，用于支撑报道的事实基础。

(6) 时间序列信息：某些新闻报道可能包含时间线或时间顺序的数据，这对于呈现事件的发展过程非常重要。

(7) 用户信息：在某些情况下，用户生成的内容（如社交媒体帖子）也被用作新闻来源，这反映了公众的观点和反应。

(8) 地理信息：新闻事件通常发生在特定的地理位置，地理信息（如地图、位置标签等）有助于读者更好地理解事件的空间背景。

每种数据类型都提供了新闻事件的不同维度的信息，多模态数据融合技术就是将这些不同来源、不同形式的数据进行有机融合，从而提取出更丰富、更全面的信息。通过将视觉信息、文本信息、音频信息、符号信息、数值信息、时间序列信息、用户信息和地理信息等多种数据类型融合在一起，智能多语种新闻采编系统能够更全面地了解新闻事件的各个方面，提升新闻报道的准确性和深度<sup>[3]</sup>。

在智能多语种新闻采编系统中，多模态数据融合技术的应用可以帮助系统更准确地捕捉新闻事件的关键信息，提高新闻采编的效率和准确性。通过对多模态数据进行融合，系统可以生成更丰富的新闻内容，为读者提供更全面、更生动的新闻报道。同时，多模态数据融合技术还可以帮助系统更好地理解用户的阅读偏好和需求，从而提供个性化的新闻推荐和定制化服务。

多模态数据融合技术也面临着一些挑战，多模态数据之间的异构性和复杂性使得数据融合过程变得复杂而繁琐，不同模态数据之间的语义鸿沟也需要被克服，以确保不同模态的信息能够准确地进行融合和解释。

## 2 多模态数据融合在新闻采编的应用

多模态数据融合在新闻采编中的应用将不断深化，为新闻行业带来革命性的变革。多模态数据融合在新闻采编的研究中展现出多样化的应用场景，这些应用不仅提升了新闻内容的丰富性、互动性，还增强了新闻报道的时效性、准确性和深度。

(1) 自动化新闻生成与增强：利用多模态数据融合技术，结合图像识别、语音转文字、视频摘要提取等技术，自动生成包含丰富视觉元素的新闻报道。例如，体育赛事报道可以通过融合比赛视频、实时数据统计、社交媒体上的观众反应等多源信息，自动生成包含图文、视频片段的综合新闻内容<sup>[4]</sup>。

(2) 事实核查与假新闻检测：多模态数据融合在新闻真实性的验证中扮演重要角色。通过分析新闻报道中的文本、图片、视频是否一致，结合地理信息、时间戳等元数据，可以有效识别虚假信息，提高新闻的可信度。

(3) 情感分析与舆论监控：在新闻采编中，情感分析与舆论监控是非常重要的环节。通过融合文本、音频和视频等多种模态的数据，可以更准确地分析出新闻中的情感倾向和舆论态势。例如，系统可以分析出新闻中人物的情感状态、公众对某个事件的态度等，为新闻编辑提供更深入、更全面的报道素材<sup>[5]</sup>。

(4) 个性化新闻推荐：随着用户需求的个性化趋势日益明显，个性化新闻推荐成为新闻采编的重要应用之一。通过融合用户的浏览历史、兴趣爱好、社交关系等多种模态的数据，可以为用户推荐更符合其需求的新闻内容。这不仅可以提高用户的满意度和忠诚度，还可以为新闻媒体带来更多的流量和收益。

(5) 跨平台内容适配：利用多模态数据融合，新闻内容可以根据不同的传播平台（如网站、移动应用、社交媒体）和设备类型（手机、平板、智能穿戴设备）进行自动优化和适配，确保最佳的用户体验。

(6) 新闻事件发现与趋势预测：通过融合社交媒体、新闻网站、电视广播等多种来源的数据，系统可以及时发现新闻事件，并提取出关键信息，如事件的时间、地点、涉及人物等。这有助于新闻编辑快速掌握新闻动态，为读者提供及时的新闻报道<sup>[6]</sup>。

(7) 跨语言新闻采编：通过融合不同语言的文本、音频和视频等多种模态的数据，可以实现跨语言的新闻采编和报道。这有助于新闻媒体扩大其影响力，提高其传播力和竞争力。

(8) 多媒体新闻报道：将文字新闻与相关的图片、视频进行融合，为受众提供更生动、直观的新闻体验。

(9) 交互式新闻体验：创建包含虚拟现实（VR）、增强现实（AR）元素的新闻报道，通过多模态数据融合技术使用户能够沉浸在新闻场景中，提供更加沉浸式和互动的新闻体验<sup>[7]</sup>。

(10) 事件还原与可视化：综合运用文本描述、现场图片、监控视频等，更全面地还原新闻事件，通过可视化手段呈现事件发展脉络。

(11) 数据新闻：把统计图表、动态图像与文字分析相结合，更清晰地展示数据背后的新闻故事。

(12) 跨媒体新闻整合：把不同媒体来源的新闻内容，如报纸、电视、网络等，依据多模态特征进行有效整合和梳理。

(13) 多模态内容摘要：通过整合文本、图像、音频等多种信息源，生成更为准确和全面的内容摘要。这有助于快速把握新闻的核心要点，提高阅读效率。

(14) 视频片段检索：在新闻报道中，可以通过多模态融合方法对视频资料进行更高效的检索和管理，快速找到所需的片段或场景<sup>[8]</sup>。

(15) 协同学习：通过多模态数据的协同学习，可以提高模型对于新闻内容的理解和处理能力，从而提升新闻编辑和推荐的相关性和准确性<sup>[9]</sup>。

(16) 跨模态检索：利用多模态融合技术实现跨模态检索，如根据文本内容自动匹配相关图片或视频，增强新闻报道的丰富性和互动性<sup>[10]</sup>。

### 3 部分关键技术

要实现上述应用，需要用到图像识别技术、语音识别与合成技术、视频分析技术、计算机视觉（CV）、虚拟现实与增强现实技术、自然语言处理技术、机器学习与深度学习技术、数据挖掘技术、信息检索与推荐技术、知识图谱技术、地理信息系统（GIS）与时空数据分析等。综合这些技术，还需要完成以下工作。

(1) 数据预处理与特征提取是多模态数据分析的关键步骤。多模态数据包括多种类型的数据，这些数据在格式、分辨率和采样率等方面存在较大差异。因此，在进行深入的多模态分析之前，需要对数据进行预处理，以统一数据格式，并提取关键特征，从而更好地捕捉数据中的关键信息。预处

理的主要目的是简化数据，使其更容易进行分析和处理。对于文本数据，预处理包括分词、词向量化等操作。分词是将文本划分为有意义的词汇单元，以便于进一步分析。词向量化是将词汇转换为数值向量，以便于计算机处理。这种向量化方法可以充分利用现有的词嵌入技术，如 Word2Vec 和 GloVe 等，将词汇表示为高维空间中的向量。这些向量可以用于计算文本间的相似性、计算文本与其他类型数据（如图像和视频）的关联性等。对于图像数据，预处理主要包括缩放、滤波等操作。缩放是将图像调整为统一的大小，以便于后续分析。滤波则是去除图像中的噪声和无关信息，从而提高图像质量。常用的滤波方法包括高斯滤波、双边滤波等。此外，还可以采用直方图均衡化、对比度增强等方法增强图像的视觉特性。视频数据的主要预处理操作有关键帧提取。关键帧提取是从视频中选取具有代表性的帧，以减少数据量并保留关键信息。常用的关键帧提取方法有基于内容的策略，如提取包含重要事件或信息的帧；基于视觉的方法，如采用运动估计、光流法等计算相邻帧之间的差异，并选取差异较大的帧作为关键帧。在完成数据预处理后，我们需要对数据进行特征提取。特征提取的目的是从原始数据中提取有价值的特征，以便于后续的分析应用。对于文本数据，可以采用词向量、TF-IDF 等方法提取特征。对于图像数据，可以采用局部特征描述子、全局特征如颜色、纹理、形状等提取特征。对于视频数据，可以采用时空特征描述子、运动轨迹等方法提取特征。

(2) 模态对齐与融合：在多元信息处理领域，不同模态的数据往往描述的是同一对象或事件的不同侧面。为了确保信息的一致性和准确性，我们需要对这些数据进行模态对齐。所谓模态对齐，就是将来自不同来源、以不同形式表示的数据转化为相同的形式，以便于进一步处理。在完成模态对齐之后，我们需要采用融合策略将这些不同模态的特征进行整合。融合策略的选择取决于数据的类型和目标任务。常见的融合方法包括早期融合、中期融合和晚期融合。早期融合主要是在特征层面将不同模态的数据进行拼接；中期融合则是在模型层面对不同模态的信息进行整合；而晚期融合则是在输出层面对融合后的特征进行处理。在新闻采编任务中，将预处理后的数据转化为相同的形式，如向量表示。这可以通过使用词嵌入技术、卷积神经网络（CNN）和循环神经网络（RNN）等方法实现。在完成模态对齐后，对各个模态的数据提取特征。这可以通过使用预训练的深度学习模型，如 BERT、CNN 和 RNN 等，提取文本、图像和音频等模态的特征。根据任务需求选择合适的融合策略，将不同模态的特征进行融合。例如，在新闻分类任务中，可

以使用早期融合方法将文本和图像的特征拼接，然后输入到神经网络中进行分类；而在新闻摘要生成任务中，可以采用晚期融合方法，先分别对文本和图像信息进行处理，再将处理后的结果进行融合，生成摘要。在完成模态对齐与融合后，可以将统一的表示输入到相应的新闻采编任务中，如新闻分类、情感分析、摘要生成等。模态对齐与融合在多元信息处理领域具有重要意义。通过对不同模态的数据进行有效整合，可以提高新闻采编任务的准确性和效果。在实际应用中，我们需要根据任务需求和数据特点选择合适的模态对齐和融合策略，以实现最佳性能。

（3）注意力机制：注意力机制是一种自动调整不同模态数据在融合过程中重要性的方法，这种机制模仿了人类在处理多模态信号时的自然能力，即使在某些模态数据质量不高的情况下，也能有效地感知环境，提高融合效果。这种机制通过自动调整不同数据源在融合过程中的权重，使得关键信息在融合过程中得到更高的重视。这一过程可以借助神经网络、循环神经网络（RNN）等先进技术实现。注意力机制的优势在于它能自动调整不同数据源的重要性，适应各种复杂场景。即使在某些数据质量不高的情况下，注意力机制仍然能够有效地识别和处理关键信息，提高整体性能。此外，注意力机制具有较强的泛化能力，可以应用于不同领域和任务。注意力机制可能导致计算复杂度的增加，尤其是在处理大规模数据时。注意力机制的透明度较低，难以解释模型的决策过程。

（4）模型训练与优化：利用大规模多模态数据集对模型进行训练，以提高模型对新闻内容的理解和处理能力。同时，通过优化算法和模型结构，提高模型的性能和稳定性。

（5）人机交互与界面设计：为了让用户能够方便地使用多模态数据融合技术，需要设计直观、易用的交互界面。这包括提供丰富的交互方式、定制化的新闻展示格式以及个性化推荐功能等。

（6）安全与隐私保护：在多模态数据融合过程中，需要关注数据的安全性和隐私保护问题。通过采用加密技术、访问控制策略以及隐私保护算法等手段，确保用户数据的安全性和隐私性。

针对多模态数据，有必要运用一系列核心技术与工具，例如数据标注及其相关工具、多模态数据的管理与存储技术等。

#### 4 未来可能的应用

在未来，多模态数据融合在新闻采编领域的应用将更加广泛和深入，以下是一些潜在的应用场景。

（1）更智能的自动化采编：系统能够更加智能地理解并融合多模态数据，自动完成从素材收集到稿件生成的全过程，且质量更高。

（2）基于脑机接口的新闻体验：通过脑机接口技术，直接根据受众的脑电波等信息来呈现个性化的多模态新闻内容。脑机接口可以根据受众的喜好，调整新闻呈现方式，如字体大小、颜色、排版等，让阅读更加舒适。通过分析受众的脑电波数据，实现精准内容推荐，满足不同人群的需求<sup>[11]</sup>。

（3）跨时空新闻场景重建：精准还原历史或远方的新闻场景，不受地域和时间的限制，让受众身临其境地感受遥远新闻事件。

（4）情绪响应新闻：利用高级情感分析技术，根据受众的情绪反馈实时调整新闻内容，如改变叙事节奏、选择更合适的视觉元素等，以最大化观众的共鸣和参与感。

（5）量子计算助力多模态融合：多模态数据作为大数据的一种重要形式，涵盖了多种类型的数据，传统的计算方法在面对如此庞大的数据量时，往往显得力不从心。量子计算最显著的特点是并行性和量子纠缠。与传统计算机相比，量子计算机在处理复杂问题时具有更高的效率。

（6）全球实时多模态新闻协作：不同地区的新闻机构能够实时共享和融合多模态资源，进行全球性的新闻协作报道。

（7）与物联网深度结合：从物联网设备中获取更广泛的多模态信息来丰富新闻内容。

（8）超高精度的虚拟主播：利用多模态数据融合打造与真人无异的虚拟主播来进行新闻播报和互动<sup>[12]</sup>。

总之，多模态数据融合在新闻采编领域的应用将不断推动新闻业的创新和发展，为读者带来更加丰富、生动、个性化的新闻体验。

#### 5 结语

多模态数据融合技术为新闻采编带来了革命性的变革，不仅提高了新闻内容的丰富性和互动性，还增强了新闻报道的时效性、准确性和深度。随着技术的不断进步和应用场景的不断拓展，多模态数据融合将在新闻采编领域发挥更加重要的作用，推动新闻媒体行业的创新和发展。

然而，多模态数据融合技术的应用也面临一些挑战和问题。例如，不同模态的数据可能存在语义差异和噪声，需要进行有效的数据清洗和预处理。同时，多模态数据的融合算法也需要不断优化和改进，以提高其性能和效率。此外，还需要考虑如何保护用户隐私和数据安全，避免数据泄露和滥用等问题。

因此，在推进多模态数据融合技术在新闻采编中的应用时，需要综合考虑技术、法律、伦理等多方面的因素，确保技术的合规性和可持续性。同时，也需要加强技术研发和人才培养，为新闻媒体的数字化转型和创新发展提供有力支撑。

参考文献:

[1] 魏丹阳.生成式人工智能在新闻采编工作中的应用与挑战[J].新闻文化建设,2024(3):94-96.

[2] 高慧琳.基于麦克卢汉媒介观的新媒介技术哲学研究[D].大连:大连理工大学,2022.

[3] 刘晓倩,张英俊,秦家虎,等.模糊认知图学习算法及应用综述[J].自动化学报,2024,50(3):450-474.

[4] 徐琦,赵子忠.中国智能媒体生态结构、应用创新与关键趋势[J].新闻与写作,2020(8): 51-58.

[5] 王宝莹.新媒体视域下的高校舆情传播研究[D].哈尔滨:黑龙江大学,2024.

[6] 郑文锋.全媒体传播时代新闻策划理念变迁与创新路径[J].传播与版权,2024(10):1-4.

[7] 许娜.人工智能技术在新闻制作过程的应用[J].电视技术,2024,48(3):103-105.

[8] 闵媛春,杨明义.我国 AI 主播研究的可视化图谱分析[J].

(上接第 201 页)

[3] 邢豪.基于时空特征的深度伪造视频篡改检测[D].太原:太原理工大学,2021.

[4] 王任颖,储贝林,杨震,等.视觉深度伪造检测技术综述[J].中国图象图形学报,2022(1):43-62.

[5] 李旭嵘,纪守领,吴春明,等.深度伪造与检测技术综述[J].软件学报,2021,32(2):496-518.

[6] 李颖,边山,王春桃,等.CNN 结合 Transformer 的深度伪造高效检测[J].中国图象图形学报,2023,28(3):804-819.

[7] 石达,芦天亮,杜彦辉,等.基于改进 CycleGAN 的人脸性别伪造图像生成模型[J].计算机科学,2022,49(2):31-39.

[8] 张小娜.基于 Java 语言的伪造图像识别检测算法[J].单片机与嵌入式系统应用,2021,21(10):49-53.

[9] RÖSSLER A, COZZOLINO D, VERDOLIVA L, et al. Face forensics ++: learning to detect manipulated facial images[C]// Proceedings of 2019 IEEE / CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 1-11.

[10] Conditional Generative Adversarial Nets (CGAN)

传媒论坛,2024,7(9):71-74.

[9] 孙立峰,宋新航,蒋树强,等.多模态协同感知与融合技术专题前言[J].软件学报,2024,35(5):2099-2100.

[10] 段毛毛,魏斌伟.基于多模态交互网络的图像描述[J].计算机技术与发展,2024,34(5):44-51.

[11] 刁生富,陈惠.元宇宙背景下虚拟数字人对人类的影响探究[J].佛山科学技术学院学报(社会科学版),2024,42(3): 31-37.

[12] 胡新华.虚拟数字人技术在新闻传播中的应用研究[J].科技传播,2024,16(3):1-3+7.

【作者简介】

李满江(1970—),男,河北蠡县人,本科,高级工程师,研究方向:中文信息处理。

鞠传森(1981—),男,山东济南人,硕士研究生,正高级工程师,研究方向:大数据应用、媒体融合。

任鹏(1971—),男,山东高密人,本科,工程师,研究方向:计算机应用。

(收稿日期:2024-06-14)

[EB/OL].(2020-07-20)[2024-01-25].[https://blog.csdn.net/qq\\_40128284/article/details/107458339](https://blog.csdn.net/qq_40128284/article/details/107458339).

[11] Keep\_Trying\_Go. WGAN 基本原理及 Pytorch 实现 WGAN [EB/OL]. (2023-05-03)[2024-01-26].[https://blog.csdn.net/Keep\\_Trying\\_Go/article/details/130471766](https://blog.csdn.net/Keep_Trying_Go/article/details/130471766).

【作者简介】

张梦萍(1989—),女,四川宜宾人,硕士研究生,助教,研究方向:计算机应用、教育技术学、现代远程教育。

牟熠(2004—),女,四川宜宾人,本科在读,研究方向:计算机科学与技术。

曾敏(2004—),女,四川内江人,本科在读,研究方向:计算机科学与技术。

肖尧馨(2004—),女,四川资阳人,本科在读,研究方向:计算机科学与技术。

(收稿日期:2024-06-12)